MATH5635 Stochastic Calculus for Finance I Autumn 2025

Chapter 5: Stochastic Calculus

Lecturer: Kenneth Ng

Preview

This chapter introduces Itô integrals, one of the fundamental concepts in stochastic calculus. We begin with the L^2 construction of Itô integrals with Brownian motion as the integrator, followed by a discussion of their key properties and computational rules. A crucial distinction from classical calculus lies in the finite, non-vanishing quadratic variation of Brownian motion, which gives rise to an additional term in Itô's lemma – the chain rule of Itô's calculus. We conclude with illustrative examples demonstrating applications of Itô's lemma.

Key topics in this chapter:

1. Constructions of Itô integrals;
2. Diffusion processes and distributions;
3. Itô's lemma;
4. Applications of Itô's lemma.

1 Constructions of Itô Integrals

The most fundamental application of ordinary calculus is to describe rates of change in modeling real-world phenomena, for example, population growth, motions, or chemical reactions. However, ordinary calculus falls short when modeling systems influenced by randomness, such as financial markets. This motivates the use of stochastic calculus, which extends the tools of calculus to accommodate randomness.

Stochastic calculus originated in the mid-20th century, primarily developed by Kiyoshi Itô, who introduced Itô integral and Itô's lemma in the 1940s. Let $X \in L^2(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a square-integrable process and B be a standard (one-dimensional) Brownian motion. We want to define the ***Itô integral*** (more generally, a ***stochastic integral***) of the form

$$I_T(X) := \int_0^T X_t \, dB_t. \tag{1}$$

One tempting approach is defined it in a *pathwise sense*, i.e., by setting

$$I_T(X)(\omega) = \int_0^T X_t(\omega)\, dB_t(\omega),\ \omega \in \Omega.$$

However, as we have learnt in the last chapter, B_t is nowhere differentiable a.s., meaning that $dB_t(\omega)$, and thus the above pathwise integral being ill-defined. Therefore, this section is devoted to giving a proper definition of (1) , specifically in the L^2-sense.

The procedures of constructing Itô integrals largely resemble the construction of Lebesgue integrals. Herein, we follow a 3-step approach for the classes of adapted processes X listed below:

1. X is a simple process;
2. X is a bounded process;
3. X is a square-integrable, adapted process.

1.1 Simple Processes

Fix $T > 0$. We first construct Itô integrals for simple processes.

Definition 1.1 (Simple processes) An adapted process $\{X_t\}_{t\in[0,T]}$ in the probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ is said to be a ***simple process*** if it is piecewise constant on $[0, T]$. To be exact, there exists a partition $\Pi = \{t_0, \ldots, t_m\}$ of $[0, T]$ with $t_0 = 0$, $t_m = T$ such that

$$X_t = \xi_0 \mathbb{1}_{\{0\}}(t) + \sum_{k=0}^{m-1} \xi_k \mathbb{1}_{(t_k,t_{k+1}]}(t), \tag{2}$$

where ξ_k is $\mathcal{F}_{t_k}$-measurable, and $\sup_{k=0,\ldots,m-1} |\xi_k| \leq C$ a.s. for some finite constant $C > 0$

Remark 1.1. A simple process is essentially ***predictable***, i.e., the value of the process at time t is known immediately before that time. A ***predictable process*** is also adapted, but the converse is not true in general. From (2), the value $X_{t_k+\delta}$ is known at t_k for any sufficiently small $\delta > 0$. The theory of Itô calculus presented below is thus built on predictable processes. Nevertheless, an adapted process with continuous sample paths (i.e., no jump) is predictable. Since we will mostly consider continuous processes in our course, we shall not distinguish the two notions in the rest of the discussions.

The Itô integral for a simple process is defined as follows.

Definition 1.2 (Itô integral of a simple process) Let X be a simple process with the representation (2). For $t \in [0,T]$, the Itô integral of X is defined as

$$I_t(X) = \int_0^t X_s \, dB_s := \sum_{k=0}^{m-1} \xi_k \big(B_{t_{k+1} \wedge t} - B_{t_k \wedge t} \big). \tag{3}$$

Equivalently, if n is the index such that $t_n < t \le t_{n+1}$, then

$$\int_0^t X_s \, dB_s = \sum_{k=0}^{m-1} \xi_k \big(B_{t_{k+1} \wedge t} - B_{t_k} \big) = \sum_{k=0}^{n-1} \xi_k \big(B_{t_{k+1}} - B_{t_k} \big) + \xi_n \big(B_t - B_{t_n} \big).$$

By definition, it is clear that the process $\{I_t(X)\}_{t\in[0,T]}$ is $\{\mathcal{F}_t\}_{t\in[0,T]}$-adapted. The following presents the major properties of Itô integrals.

Theorem 1.2 Let X be a simple process and $t \in [0,T]$. We have the following properties concerning the Itô integral $I_t(X)$:

1. (**Adaptability**) $\{I_t(X)\}_{t\in[0,T]}$ is $\{\mathcal{F}_t\}_{t\in[0,T]}$-adapted;
2. (**Continuity**) the sample paths $t \mapsto I_t(X)$ is continuous a.s.;
3. (**Linearity**) for any simple process Y and $\alpha, \beta \in \mathbb{R}$,
$$I_t(\alpha X + \beta Y) = \alpha I_t(X) + \beta I_t(Y);$$
4. (**Martingale**) $\{I_t(X)\}_{t\in[0,T]}$ is a square-integrable martingale with respect to $\{\mathcal{F}_t\}_{t\in[0,T]}$, in particular, $\mathbb{E}[I_t(X)] = 0$ for all $t \in [0,T]$;
5. (**Itô's isometry**)
$$\mathbb{E}[I_t^2(X)] = \mathbb{E}\left[\int_0^t X_s^2 \, ds \right];$$
6. (**Quadratic variation**)
$$\langle I(X) \rangle_t = \int_0^t X_s^2 \, ds.$$

Proof. The adaptability is clear from the definition (3). The continuity of the sample paths follows from the continuity of $t \mapsto B_{t_{k+1}\wedge t} - B_{t_k \wedge t}$ and (3), thanks to the continuity of the sample paths of Brownian motions.

We prove the remaining properties below.

3. Let $\Pi_1 = \{t_0, \dots, t_m\}$ and $\Pi_2 = \{s_0, \dots, s_n\}$ be two partitions of $[0,T]$ such that

$$X_t = \xi_0 \mathbb{1}_{\{0\}}(t) + \sum_{k=0}^{m-1} \xi_k \mathbb{1}_{(t_k, t_{k+1}]}(t) \text{ and } Y_t = \zeta_0 \mathbb{1}_{\{0\}}(t) + \sum_{k=0}^{n-1} \zeta_k \mathbb{1}_{(s_k, s_{k+1}]}(t).$$

Then, we can define a finer and common partition $\Pi := \Pi_1 \cup \Pi_2 = \{r_0, \dots, r_l\}$, where $l \leq m + n$. Then, X and Y can be rewritten in terms of Π as

$$X_t = \sum_{k=0}^{\ell-1} X_{r_k} \, \mathbb{1}_{(r_k, r_{k+1}]}(t), \quad Y_t = \sum_{k=0}^{\ell-1} Y_{r_k} \, \mathbb{1}_{(r_k, r_{k+1}]}(t).$$

Let $\alpha, \beta \in \mathbb{R}$. By definition of the Itô integral for simple processes,

$$I_t(\alpha X + \beta Y) = \sum_{k=0}^{\ell-1} (\alpha X_{r_k} + \beta Y_{r_k}) \left(B_{r_{k+1} \wedge t} - B_{r_k \wedge t} \right).$$

Distributing the sum, we obtain

$$\begin{aligned} I_t(\alpha X + \beta Y) &= \alpha \sum_{k=0}^{\ell-1} X_{r_k} \left(B_{r_{k+1} \wedge t} - B_{r_k \wedge t} \right) + \beta \sum_{k=0}^{\ell-1} Y_{r_k} \left(B_{r_{k+1} \wedge t} - B_{r_k \wedge t} \right) \\ &= \alpha I_t(X) + \beta I_t(Y). \end{aligned}$$

4. By the first property, the process $\{I_t(X)\}_{t \in [0,T]}$ is adapted. The integrability is also clear: for any $t \in [0, T]$, using the fact that $|\xi_k| \leq C$ a.s. and (8) below, we have

$$\mathbb{E}[|I_t(X)|^2] = \sum_{k=0}^{n-1} \mathbb{E}[\xi_k^2](t_{k+1} - t_k) + \mathbb{E}[\xi_n^2](t - t_n) \leq C^2 \sum_{k=0}^{n-1} (t_{k+1} - t_k) + C^2(t - t_n) = C^2 t,$$

where n is the integer such that $t_n < t \leq t_{n+1}$.

Finally, we verify the martingale condition. For any $s, t \in [0, T]$ with $s \leq t$, let $l \leq n$ be the unique integers such that $t_l < s \leq t_{l+1}$ and $t_n < t \leq t_{n+1}$. If $n < l$,

$$\begin{aligned} \mathbb{E}\left[I_t(X) | \mathcal{F}_s \right] &= \mathbb{E}\left[\sum_{k=0}^{n-1} \xi_k \left(B_{t_{k+1}} - B_{t_k} \right) + \xi_n (B_t - B_{t_n}) \middle| \mathcal{F}_s \right] \\ &= \mathbb{E}\left[\sum_{k=0}^{l-1} \xi_k \left(B_{t_{k+1}} - B_{t_k} \right) + \xi_l \left(B_s - B_{t_l} \right) \middle| \mathcal{F}_s \right] \\ &\quad + \mathbb{E}\left[\xi_l (B_{t_{l+1}} - B_s) + \sum_{k=l+1}^{n-1} \xi_k (B_{t_{k+1}} - B_{t_k}) + \xi_n (B_t - B_{t_n}) \middle| \mathcal{F}_s \right], \end{aligned}$$

For any $k \leq l - 1$, $t_k \leq t_{k+1} < s \leq t$, so that $\xi_k(B_{t_{k+1}} - B_{t_k})$ is $\mathcal{F}_s$-measurable. Hence,

$$\mathbb{E}\left[\sum_{k=0}^{l-1} \xi_k \left(B_{t_{k+1}} - B_{t_k} \right) + \xi_l (B_s - B_{t_l}) \middle| \mathcal{F}_s \right] = \sum_{k=0}^{l-1} \xi_k \left(B_{t_{k+1}} - B_{t_k} \right) + \xi_l (B_s - B_{t_l})$$

$$= I_s(X).$$

Next, for $k > l$, $s \le t_k$, so that $\mathcal{F}_s \subseteq \mathcal{F}_{t_k}$ Hence, by the tower property of conditional expectations,

$$\begin{aligned}
&\mathbb{E}\left[\xi_l(B_{t_{l+1}} - B_s) + \sum_{k=l+1}^{n-1} \xi_k(B_{t_{k+1}} - B_{t_k}) + \xi_n(B_t - B_{t_n})\big|\mathcal{F}_s\right] \\
&= \xi_l \mathbb{E}\left[B_{t_{l+1}} - B_s|\mathcal{F}_s\right] + \sum_{k=l+1}^{m-1} \mathbb{E}\left[\xi_k(B_{t_{k+1}} - B_{t_k})\big|\mathcal{F}_s\right] + \mathbb{E}\left[\xi_n(B_t - B_{t_n})|\mathcal{F}_s\right] \\
&= \sum_{k=l+1}^{n-1} \mathbb{E}\left[\mathbb{E}\left[\xi_k(B_{t_{k+1}} - B_{t_k})|\mathcal{F}_{t_k}\right]|\mathcal{F}_s\right] + \mathbb{E}\left[\mathbb{E}\left[\xi_n(B_t - B_{t_n})|\mathcal{F}_{t_n}\right]|\mathcal{F}_s\right] \\
&= \sum_{k=l+1}^{n-1} \mathbb{E}\left[\xi_k \mathbb{E}\left[(B_{t_{k+1}} - B_{t_k})|\mathcal{F}_{t_k}\right]|\mathcal{F}_s\right] + \mathbb{E}\left[\xi_n \mathbb{E}[B_t - B_{t_n}|\mathcal{F}_{t_n}]|\mathcal{F}_s\right] \\
&= 0,
\end{aligned}$$

since $B_{t_{k+1}} - B_{t_k}|\mathcal{F}_{t_k} = B_{t_{k+1}} - B_{t_k} \sim \mathcal{N}(0, t_{k+1} - t_k)$.

Combining the above, we arrive at $\mathbb{E}[I_t(X)|\mathcal{F}_s] = I_s(X)$ whenever $l < n$. If $l = n$, then

$$I_t(X) - I_s(X) = \xi_n(B_t - B_{t_n}) - \xi_n(B_s - B_{t_n}) = \xi_n(B_t - B_s).$$

Hence,

$$\mathbb{E}[I_t(X)|\mathcal{F}_s] = I_s(X) + \mathbb{E}[\xi_n(B_t - B_s)|\mathcal{F}_s] = I_s(X),$$

and we arrive at the same conclusion.

In particular, a martingale has constant expectations, so that $\mathbb{E}[I_t(X)] = \mathbb{E}[I_0(X)] = 0$ for any $t \ge 0$.

5. Let n be the unique integer such that $t_n < t \le t_{n+1}$. Consider the LHS of the formula:

$$\begin{aligned}
&\mathbb{E}[I_t^2(X)] \\
&= \mathbb{E}\left[\left(\sum_{k=0}^{n-1} \xi_k \left(B_{t_{k+1}} - B_{t_k}\right) + \xi_n(B_t - B_{t_n})\right)^2\right] \\
&= \underbrace{\mathbb{E}\left[\left(\sum_{k=0}^{n-1} \xi_k \left(B_{t_{k+1}} - B_{t_k}\right)\right)^2\right]}_{I_1} + \underbrace{2\sum_{k=0}^{n-1} \mathbb{E}\left[\xi_n \xi_k (B_t - B_{t_n})\left(B_{t_{k+1}} - B_{t_k}\right)\right]}_{I_2} \\
&\quad + \underbrace{\mathbb{E}\left[\xi_n^2 (B_t - B_{t_n})^2\right]}_{I_3}.
\end{aligned} \tag{4}$$

Using the tower property, the last term of (4) is given by

$$I_3 = \mathbb{E}\left[\xi_n^2(B_t - B_{t_n})^2\right] = \mathbb{E}\left[\xi_n^2\mathbb{E}\left[(B_t - B_{t_n})^2|\mathcal{F}_{t_n}\right]\right] = \mathbb{E}[\xi_n^2](t - t_n). \tag{5}$$

For the second term of (4), we again apply the tower property to obtain

$$\begin{aligned} I_2 &= 2\sum_{k=0}^{n-1}\mathbb{E}\left[\xi_n\xi_k(B_t - B_{t_n})\left(B_{t_{k+1}} - B_{t_k}\right)\right] \\ &= 2\sum_{k=0}^{n-1}\mathbb{E}\left[\mathbb{E}\left[\xi_n\xi_k(B_t - B_{t_n})\left(B_{t_{k+1}} - B_{t_k}\right)|\mathcal{F}_{t_n}\right]\right] \\ &= 2\sum_{k=0}^{n-1}\mathbb{E}\left[\xi_n\xi_k\left(B_{t_{k+1}} - B_{t_k}\right)\mathbb{E}\left[B_t - B_{t_n}|\mathcal{F}_{t_n}\right]\right] \\ &= 0. \end{aligned} \tag{6}$$

For the first term of (4), we expand the expression and obtain

$$\begin{aligned} I_1 &= \mathbb{E}\left[\left(\sum_{k=0}^{n-1}\xi_k\left(B_{t_{k+1}} - B_{t_k}\right) + \xi_n(B_t - B_{t_n})\right)^2\right] \\ &= \mathbb{E}\left[\sum_{k=0}^{n-1}\xi_k^2\left(B_{t_{k+1}} - B_{t_k}\right)^2\right] + 2\sum_{k<j}^{n-1}\mathbb{E}\left[\xi_k\xi_j\left(B_{t_{k+1}} - B_{t_k}\right)\left(B_{t_{j+1}} - B_{t_j}\right)\right]. \end{aligned}$$

Following the derivation of (5) by conditioning each summand with respect to $\mathcal{F}_{t_k}$ (exercise), it is easy to see that

$$\mathbb{E}\left[\sum_{k=0}^{n-1}\xi_k^2\left(B_{t_{k+1}} - B_{t_k}\right)^2\right] = \sum_{k=0}^{n-1}\mathbb{E}[\xi_k^2](t_{k+1} - t_k).$$

Likewise, following the derivation of (6) by conditioning each summand with respect to $\mathcal{F}_{t_j}$ (exercise), we have

$$2\sum_{k<j}^{n-1}\mathbb{E}\left[\xi_k\xi_j\left(B_{t_{k+1}} - B_{t_k}\right)\left(B_{t_{j+1}} - B_{t_j}\right)\right] = 0.$$

Hence,

$$I_1 = \sum_{k=0}^{n-1}\mathbb{E}[\xi_k^2](t_{k+1} - t_k). \tag{7}$$

Combining (5), (6), and (7), we obtain

$$\mathbb{E}[I_t^2(X)] = \sum_{k=0}^{n-1}\mathbb{E}[\xi_k^2](t_{k+1} - t_k) + \mathbb{E}[\xi_n^2](t - t_n). \tag{8}$$

Next, we consider the right-hand side of the isometry. Using (2), we have

$$\begin{aligned}
\int_0^t X_s^2\,ds &= \int_0^t \left(\xi_0 \mathbb{1}_{\{0\}}(s) + \sum_{k=0}^{m-1} \xi_k \mathbb{1}_{(t_k,t_{k+1}]}(s)\right)^2 ds \\
&= \int_0^t \left(\xi_0^2 \mathbb{1}_{\{0\}}(s) + \sum_{k=0}^{m-1} \xi_k^2 \mathbb{1}_{(t_k,t_{k+1}]}(s)\right) ds \\
&= \sum_{k=0}^{m-1} \xi_k^2 \int_0^t \mathbb{1}_{(t_k,t_{k+1}]}(s)\,ds \\
&= \sum_{k=0}^{n-1} \xi_k^2 (t_{k+1}-t_k) + \xi_n^2 (t-t_n).
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\int_0^t X_s^2\,ds\right] = \mathbb{E}\left[\sum_{k=0}^{n-1} \xi_k^2 (t_{k+1}-t_k) + \xi_n^2 (t-t_n)\right] = \mathbb{E}[I_t^2(X)].$$

6. Since $\{I_t(X)\}_{t\in[0,T]}$ is a square-integrable martingale, $\{I_t^2(X)\}_{t\in[0,T]}$ is a sub-martingale. We show that for any $0 \le s \le t \le T$,

$$\mathbb{E}\left[(I_t(X) - I_s(X))^2 | \mathcal{F}_s\right] = \mathbb{E}\left[\int_s^t X_u^2\,du | \mathcal{F}_s\right]. \tag{9}$$

If this is true, by the martingale property of Itô integrals, we would have

$$\begin{aligned}
\mathbb{E}\left[I_t(X)^2 - \int_0^t X_u^2\,du | \mathcal{F}_s\right] &= 2\mathbb{E}[I_t(X)I_s(X)|\mathcal{F}_s] - I_s^2(X) - \int_0^s X_u^2\,du \\
&= 2I_s^2(X) - I_s^2(X) - \int_0^s X_u^2\,du \\
&= I_s^2(X) - \int_0^s X_u^2\,du,
\end{aligned}$$

showing that $I_t(X)^2 - \int_0^t X_u^2\,du$ is a martingale. By the definition of quadratic variations, we deduce that

$$\langle I(X)\rangle_t = \int_0^t X_s^2\,ds.$$

To prove (9), let $l \le n$ be the unique integers such that $t_l < s \le t_{l+1}$ and $t_n < t \le t_{n+1}$. Then,

$$\mathbb{E}[(I_t(X) - I_s(X))^2|\mathcal{F}_s] = \mathbb{E}\left[\left(\sum_{k=l+1}^{n-1} \xi_k (B_{t_{k+1}} - B_{t_k}) + \xi_n (B_t - B_{t_n})\right)^2 |\mathcal{F}_s\right]$$

The rest of the proof is parallel to the proof of Itô's isometry: we expand the square above as in (4), and compute each of the three resulting terms. The only difference is that the index starts at $l+1$ instead of 0 in this case. We thus omit the proof.

□

1.2 Bounded Processes

X is a bounded process if there exists a constant $0 < C < \infty$ such that

$$\sup_{t\in[0,T]} |X_t| \leq C \text{ a.s.}$$

A bounded process can be approximated by simple processes in the L^2-sense:

Theorem 1.3 Let X be a bounded adapted process. Then, there exists a sequence of simple processes $\{X^{(n)}\}_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left|X_t - X_t^{(n)}\right|^2 dt\right] = 0. \tag{10}$$

Using this approximation, we define Itô integrals for bounded adapted processes as follows.

Definition 1.3 (Itô integrals for bounded processes) Let X be a bounded adapted process, and $\{X^{(n)}\}_{n=1}^{\infty}$ be a sequence of simple processes that satisfy (10). Then, the Itô integral for X, $\{I_t(X)\}_{t\in[0,T]}$, is the unique adapted and continuous process that satisfies

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left(I_t(X) - I_t(X^{(n)})\right)^2 dt\right] = 0.$$

Remark 1.4. By Itô's isometry, the sequence $\{I(X^{(n)})\}$ is a Cauchy sequence in the L^2-sense. Hence, the limit exists uniquely in L^2, and is independent of the choice of the approximating sequence.

The properties in Theorem 1.2 remain valid.

Theorem 1.5 The properties of Itô's integrals in Theorem 1.2 also hold if X is a bounded, adapted process.

1.3 Square-Integrable Processes

X is a square-integrable process if there exists a constant $0 < C < \infty$ such that

$$\mathbb{E}\left[\int_0^T X_t^2\, dt\right] \leq C,\ t \in [0, T].$$

A square-integrable process can be approximated by bounded processes (and in turn, by simple processes) in the L^2-sense:

Theorem 1.6 Let X be a square-integrable, adapted process. Then, there exists a sequence of bounded adapted processes $\{X^{(n)}\}_{n=1}^{\infty}$ such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left|X_t - X_t^{(n)}\right|^2 dt\right] = 0. \tag{11}$$

Using this approximation, we define Itô integrals for square-integrable, adapted processes as follows.

Definition 1.4 (Itô integrals for square-integrable processes) Let X be a square-integrable, adapted process, and $\{X^{(n)}\}_{n=1}^{\infty}$ be a sequence of bounded processes that satisfy (11). Then, the Itô integral for X, $\{I_t(X)\}_{t\in[0,T]}$, is the unique adapted and continuous process that satisfies

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^T \left(I_t(X) - I_t(X^{(n)})\right)^2 dt\right] = 0.$$

The properties in Theorem 1.2 remain valid.

Theorem 1.7 The properties of Itô integrals in Theorem 1.2 also hold if X is a square-integrable, adapted process.

2 Itô's Lemma

One of the most important formula in stochastic calculus is Itô's lemma, which is the chain rule for Itô process. An Itô process is defined as follows:

Definition 2.1 (Itô Process) A continuous adapted process $X = \{X_t\}_{t\in[0,T]}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ is called an Itô process if it can be written in the form

$$X_t = X_0 + \int_0^t \mu_s\, ds + \int_0^t \sigma_s\, dB_s, \tag{12}$$

where X_0 is an $\mathcal{F}_0$-measurable random variable, $\mu = \{\mu_t\}_{t\in[0,T]}$ and $\sigma = \{\sigma_t\}_{t\in[0,T]}$ are adapted processes[a] that satisfy

$$\mathbb{E}\left[\int_0^T (|\mu_s| + |\sigma_s|^2)\, ds\right] < \infty.$$

[a]More precisely, μ and σ need to be *predictable*.

A common and useful convention is to express an Itô process in differential form as

$$dX_t = \mu_t\, dt + \sigma_t\, dB_t.$$

In this notation, $\mu_t\, dt$ is called the ***drift*** term, and $\sigma_t\, dB_t$ is called the ***diffusion*** or ***volatility*** term. In general, X is a martingale only if the drift term is zero: for any $0 \le s \le t$,

$$\begin{aligned}\mathbb{E}[X_t|\mathcal{F}_s] &= X_0 + \int_0^s \mu_u\, du + \int_0^s \sigma_u\, dB_u + \mathbb{E}\left[\int_s^t \mu_u\, du + \int_s^t \sigma\, dB_u \Big| \mathcal{F}_s\right] \\ &= X_s + \mathbb{E}\left[\int_s^t \mu_u\, du \Big| \mathcal{F}_s\right] \neq X_s.\end{aligned}$$

In particular, if $\mu_t \ge 0$ (resp. $\mu_s \le 0$) for all $t \in [0, T]$, then X is a sub-martingale (resp. supermartingale).

Proposition 2.1 The quadratic variation of the Itô process X in (12) is given by

$$\langle X \rangle_t = \int_0^t \sigma_s^2\, ds.$$

Proof. Let

$$A_t := \int_0^t \mu_s\, ds \text{ and } M_t := \int_0^t \sigma_s\, dB_s,$$

so that $X_t = X_0 + A_t + M_t$. Note that both A and M are continuous process, and M is a martingale with quadratic variation given by

$$\langle M \rangle_t = \int_0^t \sigma_s^2\, ds;$$

see Property 6 of Theorem 1.2. Hence, it suffices to prove that $\langle X \rangle_t = \langle M \rangle_t$.

For any partition $\Pi = \{t_0, \dots, t_m\}$ of $[0, t]$, we have

$$\begin{aligned}V_t^{(2)}(\Pi) &= \sum_{i=0}^{m-1} \left[X_{t_{i+1}} - X_{t_i}\right]^2 \\ &= \sum_{i=0}^{m-1} \left[(A_{t_{i+1}} + M_{t_{i+1}}) - (A_{t_i} + M_{t_i})\right]^2 \\ &= \underbrace{\sum_{i=0}^{m-1} \left[A_{t_{i+1}} - A_{t_i}\right]^2}_{I_1} + \underbrace{2\sum_{i=0}^{m-1} \left[A_{t_{i+1}} - A_{t_i}\right]\left[M_{t_{i+1}} - M_{t_i}\right]}_{I_2} + \underbrace{\sum_{i=0}^{m-1} \left[M_{t_{i+1}} - M_{t_i}\right]^2}_{I_3}.\end{aligned}$$

Using the continuity of the process A, we have

$$\begin{aligned} I_1 = \sum_{i=0}^{m-1} \left[A_{t_{i+1}} - A_{t_i}\right]^2 &\le \max_{j=0,\dots,m-1} \left|A_{t_{j+1}} - A_{t_j}\right| \sum_{i=0}^{m-1} \left|A_{t_{i+1}} - A_{t_i}\right| \\ &= \max_{j=0,\dots,m-1} \left|A_{t_{j+1}} - A_{t_j}\right| \sum_{i=0}^{m-1} \left| \int_{t_i}^{t_{i+1}} \mu_s \, ds \right| \\ &= \max_{j=0,\dots,m-1} \left|A_{t_{j+1}} - A_{t_j}\right| \int_0^t |\mu_s| \, ds \to 0 \end{aligned} \tag{13}$$

a.s. and in L^1 when $\|\Pi\| \to 0$, since $\lim_{\|\Pi\|\to 0} \max_{j=0,\dots,m-1} \left|A_{t_{j+1}} - A_{t_j}\right| = 0$. Likewise, by the continuity of M,

$$\begin{aligned} I_2 = 2\sum_{i=0}^{m-1} \left[A_{t_{i+1}} - A_{t_i}\right]\left[M_{t_{i+1}} - M_{t_i}\right] &\le 2 \max_{j=0,\dots,m-1} \left|M_{t_{j+1}} - M_{t_j}\right| \sum_{i=0}^{m-1} \left|A_{t_{i+1}} - A_{t_i}\right| \\ &\le \max_{j=0,\dots,m-1} \left|M_{t_{j+1}} - M_{t_j}\right| \int_0^t |\mu_s| \, ds \xrightarrow{L^1} 0 \end{aligned} \tag{14}$$

as $\|\Pi\| \to 0$. Finally, by the definition of quadratic variations, we have the following convergence in probability and in L^1:

$$\lim_{\|\Pi\|\to 0} I_3 = \lim_{\|\Pi\|\to 0} \sum_{i=0}^{m-1} \left[M_{t_{i+1}} - M_{t_i}\right]^2 = \langle M \rangle_t. \tag{15}$$

Therefore, combining (13)-(15) yields

$$\langle X \rangle_t = \lim_{\|\Pi\|\to 0} V_t^{(2)}(\Pi) = \lim_{\|\Pi\|\to 0} (I_1 + I_2 + I_3) = \langle M \rangle_t.$$

□

Definition 2.2 (Cross Variations) Let X and Y be two adapted processes. Let $t > 0$ and $\Pi = \{t_0, \dots, t_m\}$ be a partition of $[0, t]$. Define

$$V_t(\Pi) := \sum_{i=0}^{m-1} \left(X_{t_{i+1}} - X_{t_i}\right)\left(Y_{t_{i+1}} - Y_{t_i}\right).$$

Then, the ***cross variation*** of X and Y, denoted by $\langle X, Y \rangle$, is defined as

$$\langle X, Y \rangle_t = \lim_{\|\Pi\|\to 0} V_t(\Pi).$$

Alternatively, the cross variation can also be written as

$$\langle X, Y \rangle_t = \frac{\langle X + Y \rangle_t - \langle X - Y \rangle_t}{4}. \tag{16}$$

Remark 2.2.

1. To see that the cross variation can be written as (16), note that

$$\begin{aligned}
&\sum_{i=0}^{m-1}\left[\left(X_{t_{i+1}}+Y_{t_{i+1}}\right)-\left(X_{t_i}+Y_{t_i}\right)\right]^2+\sum_{i=0}^{m-1}\left[\left(X_{t_{i+1}}-Y_{t_{i+1}}\right)-\left(X_{t_i}-Y_{t_i}\right)\right]^2\\
&=\sum_{i=0}^{m-1}\left[\left(X_{t_{i+1}}-X_{t_i}\right)^2+2\left(X_{t_{i+1}}-X_{t_i}\right)\left(Y_{t_{i+1}}-Y_{t_i}\right)+\left(Y_{t_{i+1}}-Y_{t_i}\right)^2\right]\\
&\quad-\sum_{i=0}^{m-1}\left[\left(X_{t_{i+1}}-X_{t_i}\right)^2-2\left(X_{t_{i+1}}-X_{t_i}\right)\left(Y_{t_{i+1}}-Y_{t_i}\right)+\left(Y_{t_{i+1}}-Y_{t_i}\right)^2\right]\\
&=4\sum_{i=0}^{m-1}\left(X_{t_{i+1}}-X_{t_i}\right)\left(Y_{t_{i+1}}-Y_{t_i}\right).
\end{aligned}$$

   By passing to the limit $\|\Pi\|\to 0$, we see that (16) holds.

2. If X and Y are square-integrable martingales, then the cross variation of X and Y is the unique process such that $M_t := X_tY_t - \langle X, Y\rangle_t$ is a martingale.

The cross variation of two Itô processes is given as follows.

Proposition 2.3 Let X_t and Y_t be Itô processes of the form

$$X_t = X_0 + \int_0^t \mu_s^X\,ds + \int_0^t \sigma_s^X\,dB_s, \qquad Y_t = Y_0 + \int_0^t \mu_s^Y\,ds + \int_0^t \sigma_s^Y\,dB_s,$$

where $\mu_t^X, \mu_t^Y, \sigma_t^X, \sigma_t^Y$ are adapted processes. Then,

$$\langle X, Y\rangle_t = \int_0^t \sigma_s^X\sigma_s^Y\,ds.$$

Proof. Note that

$$\begin{aligned}
X_t + Y_t &= X_0 + Y_0 + \int_0^t \left(\mu_s^X + \mu_s^Y\right)ds + \int_0^t \left(\sigma_s^X + \sigma_s^Y\right)dB_s,\\
X_t - Y_t &= X_0 - Y_0 + \int_0^t \left(\mu_s^X - \mu_s^Y\right)ds + \int_0^t \left(\sigma_s^X - \sigma_s^Y\right)dB_s.
\end{aligned}$$

By Proposition 2.1, we have

$$\langle X+Y\rangle_t = \int_0^t \left(\sigma_s^X+\sigma_s^Y\right)^2 ds \quad \text{and} \quad \langle X-Y\rangle_t = \int_0^t \left(\sigma_s^X-\sigma_s^Y\right)^2 ds.$$

Hence, using (16), we have

$$\begin{aligned}\langle X, Y\rangle_t &= \frac{\langle X+Y\rangle_t - \langle X-Y\rangle_t}{4} \\ &= \frac{1}{4}\int_0^t \left[\left(\sigma_s^X + \sigma_s^Y\right)^2 - \left(\sigma_s^X - \sigma_s^Y\right)^2\right] ds \\ &= \int_0^t \sigma_s^X \sigma_s^Y \, ds.\end{aligned}$$

□

In classical calculus, if we know the derivative of a function x_t with respect to t, we can also compute the derivative of the composite function $f(x_t)$ using chain rule. In stochastic calculus, In stochastic calculus, Itô's lemma serves as the analogue of the chain rule, providing the formula for the Itô diffusion $f(X_t)$ for an Itô process X_t.

Theorem 2.4 (Itô's Lemma) Let $f : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a function of class $\mathcal{C}^{1,2}$, i.e., $(t,x) \mapsto f(t,x)$ is continuously differentiable in t, and twice continuously differentiable in x. Let X be an Itô process given by (12). Then, for any $t \in [0,T]$,

$$\begin{aligned}f(t, X_t) &= f(0, X_0) + \int_0^t f_t(s, X_s)\, ds + \int_0^t f_x(t, X_s)\, dX_s + \int_0^t \frac{1}{2} f_{xx}(t, X_s)\, d\langle X\rangle_s \\ &= f(0, X_0) + \int_0^t \left(f_t(s, X_s) + \mu_s f_x(s, X_s) + \frac{1}{2}\sigma_s^2 f_{xx}(s, X_s)\right) ds \qquad (17) \\ &\quad + \int_0^t \mu_s f_x(s, X_s)\, dB_s.\end{aligned}$$

Equivalently, in differential form, we have

$$\begin{aligned}df(t, X_t) &= f_t(t, X_t)\, dt + f_x(t, X_t)\, dX_t + \frac{1}{2} f_{xx}(t, X_t)\, d\langle X\rangle_t \\ &= \left(f_t(t, X_t) + \mu_t f_x(t, X_t) + \frac{1}{2}\sigma_t^2 f_{xx}(t, X_t)\right) dt + \mu_t f_x(t, X_t)\, dB_t.\end{aligned}$$

Remark 2.5. In classical calculus, suppose that $x : [0,T] \to \mathbb{R}$ is a differentiable function satisfying $\frac{dx_t}{dt} = \mu_t$. By the chain rule, we have

$$\frac{df(t, x_t)}{dt} = f_t(t, x_t) + f_x(t, x_t)\frac{dx_t}{dt} = f_t(t, x_t) + f_x(t, x_t)\mu_t.$$

The additional term $\frac{1}{2} f_{xx}(t, X_t) d\langle X\rangle_t$ in Itô's lemma is due to the non-zero quadratic variation of Brownian motions.

Proof. A rigorous proof of Itô's lemma will require techniques of localization and approximations. Hence, we only outline the main idea of the proof below.

Let Π be a partition of $[0, t]$. Then, by Taylor expansion (or the mean value theorem),

$$f(t, X_t) - f(0, X_0) = \sum_{i=0}^{m-1} f_t(\eta_i, X_{\eta_i})(t_{i+1} - t_i) + \sum_{i=0}^{m-1} f_x(t_i, X_{t_i}) \left(X_{t_{i+1}} - X_{t_i}\right) + \frac{1}{2} \sum_{i=0}^{m-1} f_{xx}(t_i, \xi_i) \left[X_{t_{i+1}} - X_{t_i}\right]^2, \tag{18}$$

where for each $i = 0, \ldots, m-1$, η_i and ξ_i are $\mathcal{F}_{t_{i+1}}$-measurable, with $\eta_i \in [t_i, t_{i+1}]$, and ξ_i lying in between X_{t_i} and $X_{t_{i+1}}$.

By passing to the limit $\|\Pi\| \to 0$, one has

$$\sum_{i=0}^{m-1} f_t(\eta_i, X_{\eta_i})(t_{i+1} - t_i) \xrightarrow{L^1} \int_0^t f_t(s, X_s)\, ds,$$

$$\sum_{i=0}^{m-1} f_x(t_i, X_{t_i}) \left(X_{t_{i+1}} - X_{t_i}\right) \xrightarrow{L^1} \int_0^t f_x(t, X_s)\, dX_s = \int_0^t f_x(s, X_s)\mu_s\, ds + \int_0^t f_x(s, X_s)\sigma_s\, dB_s. \tag{19}$$

For the third summand, we consider

$$\sum_{i=0}^{m-1} f_{xx}(t_i, \xi_i) \left[X_{t_{i+1}} - X_{t_i}\right]^2 = \sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[X_{t_{i+1}} - X_{t_i}\right]^2 + \sum_{i=0}^{m-1} \left(f_{xx}(t_i, X_{t_i}) - f_{xx}(t, \xi_i)\right) \left[X_{t_{i+1}} - X_{t_i}\right]^2. \tag{20}$$

By the continuity of $f_{xx}(t, \cdot)$ and X, as $\|\Pi\| \to 0$, we have

$$\sum_{i=0}^{m-1} \left(f_{xx}(t_i, X_{t_i}) - f_{xx}(t, \xi_i)\right) \left[X_{t_{i+1}} - X_{t_i}\right]^2 \xrightarrow{L^1} 0. \tag{21}$$

On the other hand, similar to the proof of Proposition 2.1, we let $A_t := \int_0^t \mu_s\, ds$ and $M_t := \int_0^t \sigma_s\, dB_s$. We then decompose the first term of (20) into

$$\sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[X_{t_{i+1}} - X_{t_i}\right]^2$$

$$= \sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[A_{t_{i+1}} - A_{t_i}\right]^2 + 2\sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[A_{t_{i+1}} - A_{t_i}\right] \left[M_{t_{i+1}} - M_{t_i}\right]$$
$$+ \sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[M_{t_{i+1}} - M_{t_i}\right]^2 \tag{22}$$

Following the derivation of (13)-(15), one can show that, as $\|\Pi\| \to 0$,

$$\sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[A_{t_{i+1}} - A_{t_i}\right]^2 \xrightarrow{L^1} 0$$
$$2\sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[A_{t_{i+1}} - A_{t_i}\right] \left[M_{t_{i+1}} - M_{t_i}\right] \xrightarrow{L^1} 0,$$
$$\sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[M_{t_{i+1}} - M_{t_i}\right]^2 \xrightarrow{L^1} \int_0^t f_{xx}(s, X_s)\, d\langle M\rangle_s$$
$$= \int_0^t f_{xx}(s, X_s)\sigma_s^2\, ds. \tag{23}$$

Therefore, combining (20)-(22) yields

$$\sum_{i=0}^{m-1} f_{xx}(t_i, X_{t_i}) \left[X_{t_{i+1}} - X_{t_i}\right]^2 \xrightarrow{L^1} \int_0^t f_{xx}(s, X_s)\sigma_s^2\, ds. \tag{24}$$

as $\|\Pi\| \to 0$.

Finally, the formula follows by combining (18), (19), and (24)

□

3 Applications of Itô's Lemma

This section presents a number of examples of using Itô's lemma. In particular, we will see how Itô's lemma can be used to show the martingale property and compute moments of an Itô process.

Example 3.1 Using Itô's lemma, show that

$$\int_0^t B_s\, dB_s = \frac{1}{2}B_t^2 - \frac{t}{2}.$$

Solution. We let $f(x) = \frac{1}{2}x^2$. By Itô's lemma,

$$
\begin{aligned}
d\left(\frac{1}{2}B_t^2\right) = df(B_t) &= f'(B_t)\,dB_t + \frac{1}{2}f''(B_t)d\langle B\rangle_t \\
&= B_t\,dB_t + \frac{1}{2} \times 1\,dt \\
&= B_t\,dB_t + \frac{1}{2}\,dt.
\end{aligned}
$$

Integrating both sides from 0 to t, and using the fact that $B_0 = 0$, we have

$$\frac{1}{2}B_t^2 = \int_0^t B_s\,dB_s + \int_0^t \frac{1}{2}\,ds = \int_0^t B_s\,dB_s + \frac{t}{2}.$$

We thus obtain the identity be rearranging the above. □

Remark 3.1. Let f be a differentiable function with $f(0) = 0$. By the change of variables formula in classical integrations, we have

$$\int_0^t f(s)\,df(s) = \frac{1}{2}f^2(t).$$

Comparing this with Example 3.1, we see that there is an extra correction term $-t/2$ in stochastic calculus.

3.1 Product Rule

Theorem 3.2 (Product Rule) Let X_t and Y_t be Itô processes of the form

$$dX_t = \mu_t^X dt + \sigma_t^X dW_t, \qquad dY_t = \mu_t^Y dt + \sigma_t^Y dW_t,$$

where $\mu_t^X, \mu_t^Y, \sigma_t^X, \sigma_t^Y$ are adapted processes satisfying the usual integrability conditions. Then the product $Z_t := X_t Y_t$ is also an Itô process, and it satisfies

$$d(X_t Y_t) = X_t\,dY_t + Y_t\,dX_t + d\langle X, Y\rangle_t.$$

Proof. We use the identity

$$X_t Y_t = \frac{1}{4}\left((X_t + Y_t)^2 - (X_t - Y_t)^2\right).$$

Define $A_t := X_t + Y_t$ and $B_t := X_t - Y_t$. Then both A_t and B_t are Itô processes. Applying Itô's formula to A_t^2 and B_t^2, we obtain:

$$dA_t^2 = 2A_t\,dA_t + d\langle A\rangle_t, \qquad dB_t^2 = 2B_t\,dB_t + d\langle B\rangle_t.$$

Since $A_t = X_t + Y_t$, we have $dA_t = dX_t + dY_t$, and similarly $dB_t = dX_t - dY_t$. Also, the quadratic variation satisfies:

$$d\langle A\rangle_t = d\langle X\rangle_t + 2\,d\langle X,Y\rangle_t + d\langle Y\rangle_t,$$

$$d\langle B\rangle_t = d\langle X\rangle_t - 2\,d\langle X,Y\rangle_t + d\langle Y\rangle_t.$$

Thus, taking the difference,

$$d(A_t^2 - B_t^2) = 2A_t\,dA_t - 2B_t\,dB_t + (d\langle A\rangle_t - d\langle B\rangle_t)\,.$$

But observe:

$$2A_t\,dA_t - 2B_t\,dB_t = 2(X_t+Y_t)(dX_t+dY_t) - 2(X_t-Y_t)(dX_t-dY_t) = 4(X_t\,dY_t + Y_t\,dX_t),$$

and

$$d\langle A\rangle_t - d\langle B\rangle_t = 4\,d\langle X,Y\rangle_t.$$

Therefore,

$$d(A_t^2 - B_t^2) = 4(X_t\,dY_t + Y_t\,dX_t) + 4\,d\langle X,Y\rangle_t.$$

Dividing both sides by 4, we get:

$$d(X_tY_t) = X_t\,dY_t + Y_t\,dX_t + d\langle X,Y\rangle_t.$$

□

3.2 Martingale Property

One useful application of Itô's lemma is to show the (sub/super)-martingale property of an Itô process. Let X be an Itô process with the differential form

$$dX_t = \mu_t\,dt + \sigma_t\,dB_t.$$

Suppose that the drift term vanishes, i.e., $\mu_\cdot \equiv 0$. Then, $X_t = \int_0^t \sigma_s\,dB_s$ is essentially an Itô integral, which is a martingale. Similarly, if X has a non-negative (resp. non-positive) drift term, X will be a sub-martingale (resp. supermartingale).

Example 3.2 Let X be an Itô process with the representation (12). Using Itô's lemma, show that

$$M_t := X_t^2 - \int_0^t \left(2\mu_s X_s + \sigma_s^2\right)\,ds$$

is a martingale.

Solution. Let $M_t := f(X_t)$, where $f(x) = x^2$. By Itô's lemma,

$$\begin{aligned} dM_t &= \left(f_x(X_t)\mu_t + \frac{1}{2}\sigma_t^2 f_{xx}(X_t) \right) dt + f_x(X_t)\sigma_t \, dB_t \\ &= \left(2\mu_t X_t + \frac{1}{2}\sigma_t^2 \times 2 \right) dt + 2\sigma_t X_t \, dB_t \\ &= \left(2\mu_t X_t + \sigma_t^2 \right) \, dt + 2\sigma_t X_t \, dB_t. \end{aligned}$$

Integrating both sides with respect to t, we obtain

$$M_t - \int_0^t \left(2\mu_s X_s + \sigma_s^2 \right) \, ds = X_t^2 - \int_0^t \left(2\mu_s X_s + \sigma_s^2 \right) \, ds = X_0^2 + 2\int_0^t \sigma_s X_s \, dB_s.$$

Since the LHS can be represented as an Itô integral, we conclude that the process is a martingale.

□

Example 3.3 Suppose that the price of a risky asset $\{S_t\}_{t\geq 0}$ is given by

$$S_t = S_0 \exp\left(\int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2 \right) \, ds + \int_0^t \sigma_s \, dB_s \right).$$

(a) Represent S as an Itô process. Is S a martingale in general?

(b) Let $V_t := e^{-rt} S_t$. Represent V as an Itô process and show that V_t is a martingale if $\mu_\cdot \equiv r$.

Solution.

(a) Let $f(x) := e^x$, and

$$X_t := \int_0^t \left(\mu_s - \frac{1}{2}\sigma_s^2 \right) \, ds + \int_0^t \sigma_s \, dB_s.$$

By Itô's lemma,

$$\begin{aligned} dS_t = df(X_t) &= f_x(t, X_t) \, dX_t + \frac{1}{2} f_{xx}(t, X_t) d\langle X \rangle_t \\ &= S_t \left[\left(\mu_t - \frac{1}{2}\sigma_t^2 \right) + \frac{1}{2}\sigma_t^2 \right] dt + S_t \sigma_t \, dB_t \\ &= S_t \mu_t \, dt + S_t \sigma_t \, dB_t. \end{aligned}$$

Since the drift term does not vanish unless $\mu_\cdot \equiv 0$, S is in general NOT a martingale.

(b) Let $g(t,x) := e^{-rt}x$, so that $V_t = g(t, S_t)$. By applying Itô's lemma to $g(t, S_t)$ (or simply using product rule)

$$\begin{aligned} dV_t = dg(t, S_t) &= g_t(t, S_t)\, dt + g_x(t, S_t)\, dS_t + \frac{1}{2} g_{xx}(t, S_t) d\langle S\rangle_t \\ &= e^{-rt} S_t \left(-r + \mu_t\right) dt + e^{-rt} S_t \sigma_t\, dB_t \\ &= V_t(\mu_t - r) dt + V_t \sigma_t\, dB_t. \end{aligned}$$

If $\mu_\cdot \equiv r$, the expression is reduced to

$$dV_t = V_t \sigma_t\, dB_t.$$

In this case, V becomes a martingale.

□

3.3 Distributions and Computations of Expected Values

Another important use of Itô's lemma is to compute expected values, especially moments, of Itô processes. Let X be an Itô process of the form (12). We can take expectation on both sides to give

$$\mathbb{E}[X_t] = \mathbb{E}\left[\int_0^t \mu_s\, ds\right] + \mathbb{E}\left[\int_0^t \sigma_s\, dB_s\right] = \int_0^t \mathbb{E}\left[\mu_s\right] ds.$$

We can express the above in terms of a differential equation, especially when the drift term μ_t depends on X_t:

$$\frac{d}{dt}\mathbb{E}[X_t] = \mathbb{E}[\mu_t].$$

If the drift term depends on X, then $\mathbb{E}[X_t]$ is computed by solving the above differential equation.

To compute higher moments $\mathbb{E}[X_t^p]$, $p \geq 1$, we can apply Itô's lemma to X_t^p, followed by deriving the differential equation satisfied by $\mathbb{E}[X_t^p]$.

Example 3.4 Suppose that X is an ***Ornstein–Uhlenbeck (OU) process*** which satisfies the following:

$$dX_t = \theta(\mu - X_t)\, dt + \sigma\, dB_t,\ X_0 = r,$$

where $\theta, \sigma, r > 0$ and $\mu \in \mathbb{R}$. The OU process is also called the ***Vasicek model*** in interest rate models.

(a) Compute $\mathbb{E}[X_t]$.
(b) Compute $\mathrm{Var}[X_t]$.
(c) Compute $\lim_{t\to\infty} \mathbb{E}[X_t]$ and $\lim_{t\to\infty} \mathrm{Var}[X_t]$.

Solution.

(a) Taking expectation on both sides of the equation, we have

$$d\mathbb{E}[X_t] = \theta(\mu - \mathbb{E}[X_t])dt,\ \mathbb{E}[X_0] = r,$$

Equivalently, if we let $f(t) := \mathbb{E}[X_t]$, we arrive at following *first-order differential equation*:

$$f'(t) = \theta(\mu - f(t)),\ f(0) = r.$$

To solve the differential equation, we apply the method of integrating factor: by multiplying both sides with $e^{\theta t}$, we have

$$\frac{d}{dt}(e^{\theta t} f(t)) = e^{\theta t}(f'(t) + \theta f(t)) = \theta\mu e^{\theta t}.$$

Integrating both sides yields

$$e^{\theta t} f(t) - e^{\theta \times 0} f(0) = \int_0^t \theta\mu e^{\theta s}\, ds = \mu\left(e^{\theta t} - 1\right).$$

Therefore,

$$\mathbb{E}[X_t] = f(t) = re^{-\theta t} + \mu(1 - e^{-\theta t}).$$

(b) We need to compute $\mathbb{E}[X_t^2]$. To this end, by applying Itô's lemma to X_t^2, we have

$$\begin{aligned} dX_t^2 &= 2X_t\, dX_t + \frac{1}{2}(2)d\langle X\rangle_t \\ &= \left(2\theta(\mu X_t - X_t^2) + \sigma^2\right)\, dt + 2\sigma X_t\, dB_t \end{aligned}$$

Taking expectation on both sides and let $g(t) := \mathbb{E}[X_t^2]$ yields

$$g'(t) = 2\theta\mu f(t) - 2\theta g(t) + \sigma^2,\ g(0) = r^2.$$

Using the method of integrating factor,

$$\frac{d}{dt}\left(e^{2\theta t} g(t)\right) = e^{2\theta t}\left(g'(t) + 2\theta g(t)\right) = e^{2\theta t}\left(2\theta\mu f(t) + \sigma^2\right).$$

Therefore, integrating both sides yields

$$\begin{aligned} e^{2\theta t} g(t) - e^{2\theta(0)} g(0) &= \int_0^t e^{2\theta s}\left(2\theta\mu f(s) + \sigma^2\right)\, ds \\ &= \int_0^t e^{2\theta s}\left[2\theta\mu\left((r-\mu)e^{-\theta s} + \mu\right) + \sigma^2\right]\, ds \\ &= \int_0^t \left[2\theta\mu(r-\mu)e^{\theta s} + \left(\sigma^2 + 2\theta\mu^2\right) e^{2\theta s}\right]\, ds \end{aligned}$$

$$= 2\mu(r-\mu)\left(e^{\theta t}-1\right) + \frac{\sigma^2 + 2\theta\mu^2}{2\theta}\left(e^{2\theta t}-1\right)$$
$$= \left(\frac{\sigma^2}{2\theta}+\mu^2\right)e^{2\theta t} + 2\mu(r-\mu)e^{\theta t} - \left(\frac{\sigma^2}{2\theta} + 2\mu r - \mu^2\right).$$

Hence,

$$\begin{aligned} g(t) &= r^2 e^{-2\theta t} + \left(\frac{\sigma^2}{2\theta}+\mu^2\right) + 2\mu(r-\mu)e^{-\theta t} - \left(\frac{\sigma^2}{2\theta}+2\mu r-\mu^2\right)e^{-2\theta t} \\ &= \frac{\sigma^2}{2\theta} + \mu^2 + 2\mu(r-\mu)e^{-\theta t} + \left[(r-\mu)^2 - \frac{\sigma^2}{2\theta}\right]e^{-2\theta t}. \end{aligned}$$

Therefore,

$$\begin{aligned} \mathrm{Var}[X_t] &= g(t) - f^2(t) \\ &= g(t) - \left[(r-\mu)e^{-\theta t}+\mu\right]^2 \\ &= \frac{\sigma^2}{2\theta}\left(1-e^{-2\theta t}\right). \end{aligned}$$

(c) Using the formula of $\mathbb{E}[X_t]$ and $\mathrm{Var}[X_t]$,

$$\lim_{t\to\infty}\mathbb{E}[X_t] = \lim_{t\to\infty}\left[(r-\mu)e^{-\theta t}+\mu\right] = \mu,$$
$$\lim_{t\to\infty}\mathrm{Var}[X_t] = \lim_{t\to\infty}\frac{\sigma^2}{2\theta}\left(1-e^{-2\theta t}\right) = \frac{\sigma^2}{2\theta}.$$

□

We present some results that utilize Itô's lemma to characterize the distributions of processes.

Proposition 3.3 Let f be a deterministic function satisfying $\int_0^T |f(t)|^2\,dt < \infty$. Then, the Itô integral of f has the following distribution:

$$\int_0^T f(t)\,dB_t \sim \mathcal{N}\left(0, \int_0^T |f(t)|^2\,dt\right).$$

Proof. For $t \in [0,T]$, let $X_t := \int_0^t f(s)\,dB_s$ and $Y_t := e^{\lambda X_t}$, where $\lambda \in \mathbb{R}$. Applying Itô's lemma to Y_t yields

$$\begin{aligned} dY_t &= \lambda Y_t\,dX_t + \frac{\lambda^2}{2}Y_t\,d\langle X\rangle_t \\ &= \frac{\lambda^2}{2}f^2(t)Y_t\,dt + \lambda Y_t f(t)\,dB_t. \end{aligned}$$

Let $M_\lambda(t) := \mathbb{E}[Y_t] = \mathbb{E}[e^{\lambda X_t}]$, so that $\lambda \mapsto M_\lambda(t)$ is the mgf of X_t. Using the equation for Y_t, we have

$$M_\lambda'(t) = \frac{\lambda^2}{2} f^2(t) M_\lambda(t).$$

By multiplying the differential equation with the integrating factor $e^{-\frac{\lambda^2}{2}\int_0^t f^2(s)\,ds}$, we have

$$\frac{d}{dt}\left(e^{-\frac{\lambda^2}{2}\int_0^t f^2(s)\,ds} M_\lambda(t)\right) = e^{-\frac{\lambda^2}{2}\int_0^t f^2(s)\,ds}\left(M_\lambda'(t) - \frac{\lambda^2}{2} f^2(s) M_\lambda(t)\right) = 0.$$

Hence,

$$M_\lambda(T) = e^{\frac{\lambda^2}{2}\int_0^T f^2(t)\,dt} M_\lambda(0) = e^{\frac{\lambda^2}{2}\int_0^T f^2(t)\,dt},$$

since $M_\lambda(0) = \mathbb{E}[e^{\lambda X_0}] = \mathbb{E}[e^0] = 1$. Note that $e^{\frac{\lambda^2}{2}\int_0^T f^2(t)\,dt}$ is the mgf of the distribution $\mathcal{N}(0, \int_0^T f^2(t)\,dt)$. Therefore, the proof is complete. □

In the last chapter, we know that the quadratic variation of a Brownian motion is t. The following result, proven by Itô's lemma, shows the converse.

Theorem 3.4 (Lévy's Characterization) Let $\{X_t\}_{t\geq 0}$ be an adapted process with $X_0 = 0$. Suppose that:

1. X is a martingale with continuous sample paths;
2. $\langle X \rangle_t = t$ for all $t \geq 0$.

Then X is a standard Brownian motion.

Proof. Fix $\lambda \in \mathbb{R}$, define the process Y_t by $Y_t := e^{\lambda X_t}$. By Itô's lemma and the fact that $\langle X \rangle_t = t$, we have

$$dY_t = \lambda Y_t\,dX_t + \frac{\lambda^2}{2} Y_t\,d\langle X \rangle_t = \lambda Y_t\,dX_t + \frac{\lambda^2}{2} Y_t\,dt.$$

Integrating both sides with respect to t, we have

$$Y_t = Y_0 + \lambda \int_0^t Y_s\,dX_s + \frac{\lambda^2}{2}\int_0^t Y_s\,ds.$$

Since X_t is a martingale, the integral $\int_0^t Y_s\,dX_s$ is also a martingale[1] Taking expectation on both sides yields

$$\mathbb{E}[Y_t] = 1 + \frac{\lambda^2}{2}\int_0^t \mathbb{E}[Y_s]\,ds,$$

[1]This can be shown by constructing a general stochastic integral using X as an integrator in place of the Brownian motion. Alternatively, we can use the ***martingale representation theorem*** to represent X_t as an Itô integral. The details will be covered in the last chapter.

which is a first-order integral equation. Applying the method of integrating factor, we have that

$$\mathbb{E}\left[e^{\lambda X_t}\right] = \mathbb{E}[Y_t] = e^{\frac{\lambda^2 t}{2}}. \tag{25}$$

Therefore, $X_t \sim \mathcal{N}(0, t)$.

This findings can be used to show that X has independent, stationary, and Gaussian increment. To this end, fix $\lambda, \mu \in \mathbb{R}$ and $s \geq 0$. For any $t \geq s$, we define the process

$$Z_t := e^{\lambda X_s + \mu(X_t - X_s)} = Y_s e^{\mu(X_t - X_s)}.$$

By treating s as fixed and applying Itô's lemma to Z_t with respect to t, and using the quadratic variation of X, we have

$$dZ_t = \mu Z_t \, dX_t + \frac{\mu^2}{2} Z_t \, dt.$$

Integrating both sides with from s to t, followed by taking expectations, we have

$$\mathbb{E}[Z_t] = \mathbb{E}[Z_s] + \mathbb{E}\left[\mu \int_s^t dX_u\right] + \frac{\mu^2}{2} \int_s^t \mathbb{E}[Z_u] \, du = \mathbb{E}[Y_s] + \frac{\mu^2}{2} \int_s^t \mathbb{E}[Z_u] \, du.$$

Solving this integral equations yields

$$\mathbb{E}[e^{\lambda X_s + \mu(X_t - X_s)}] = \mathbb{E}[Z_t] = \mathbb{E}[Y_s] e^{\frac{\mu^2(t-s)}{2}} = e^{\frac{\lambda^2 s}{2} + \frac{\mu^2(t-s)}{2}},$$

where the last equality follows from (25). Since the joint mgf of $(X_s, X_t - X_s)$ into the product of the marginals, and both are Gaussian, this proves that $(X_s, X_t - X_s)$ follows a multivariate normal distribution with covariance matrix

$$\begin{pmatrix} s & 0 \\ 0 & t-s \end{pmatrix}.$$

Therefore, X has independent, stationary, and Gaussian increments as desired. □

4 Multivariate Stochastic Calculus

Itô's lemma can be generalized into the following multi-dimensional version, which can be proven using a similar argument as the one-dimensional Itô's lemma by considering the Taylor series expansion of a multivariate function.

Theorem 4.1 (Multivariate Itô's Lemma) Let $\mathbf{B}_t = (B_t^j)_{j=1}^d$ be a d-dimensional standard Brownian motion, and $\mathbf{X}_t = (X_t^i)_{i=1}^n$ be an n-dimensional Itô process with the expression

$$dX_t^i = \mu_t^i \, dt + \sum_{j=1}^{d} \sigma_t^{i,j} \, dB_t^j.$$

Then, for any $f : [0, \infty) \times \mathbb{R}^n \to \mathbb{R}$ that is of $\mathcal{C}^{1,2}$ (i.e., continuously differentiable in the t-variable, and twice continuously differentiable in the x-variable),

$$\begin{aligned} df(t, X_t^1, \ldots, X_t^n) &= f_t(t, X_t^1, \ldots, X_t^n) \, dt + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(t, X_t^1, \ldots, X_t^n) \, d\langle X^i \rangle_t \\ &\quad + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_k}(t, X_t^1, \ldots, X_t^n) \, d\langle X^i, X^k \rangle_t \\ &= \left[f_t(t, X_t^1, \ldots, X_t^n) + \sum_{i=1}^{n} \mu_t^i \frac{\partial f}{\partial x_i}(t, X_t^1, \ldots, X_t^n) \right. \\ &\quad \left. + \frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{n} \sum_{j=1}^{d} \frac{\partial^2 f}{\partial x_i \partial x_k}(t, X_t^1, \ldots, X_t^n) \sigma_t^{i,j} \sigma_t^{k,j} \right] dt \\ &\quad + \sum_{i=1}^{n} \sum_{j=1}^{d} \sigma_t^{i,j} \frac{\partial f}{\partial x_i}(t, X_t^1, \ldots, X_t^n) \, dB_t^j. \end{aligned}$$

The product rule presented in Proposition Theorem 3.2 can be proven by applying the multivariate Itô's lemma to $f(x, y) = xy$. We also have the following generalization of Lévy's characterization theorem (Theorem 3.4).

Theorem 4.2 (Multi-dimensional Lévy's Characterization) Let $\mathbf{X} = (X_t^1, \ldots, X_t^n)_{t \geq 0}$ be an $\mathbb{R}^n$-valued adapted process with $X_0 = 0$. Suppose that:

1. Each component X^i is a martingale with continuous sample paths;
2. For all $i, j = 1, \ldots, n$, the cross-variation satisfies

$$\langle X^i, X^j \rangle_t = \delta_{ij} t = \begin{cases} t, & \text{if } i = j; \\ 0, & \text{if } i \neq j \end{cases} \quad \text{for all } t \geq 0;$$

Then X is a standard n-dimensional Brownian motion.

Proof. Fix $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and define the process $Y_t := f(t, X_t^1, \ldots, X_t^n)$, where

$$f(t, x_1, \ldots, x_n) := \exp\left(\sum_{i=1}^{n} \lambda_i x_i \right).$$

By the multivariate Itô's lemma, we have

$$\begin{aligned}
dY_t &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(t, X_t)\, dX_t^i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X_t)\, d\langle X^i, X^j\rangle_t \\
&= \sum_{i=1}^{n} \lambda_i \exp\left(\sum_{k=1}^{n} \lambda_k X_t^k\right) dX_t^i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \exp\left(\sum_{k=1}^{n} \lambda_k X_t^k\right) d\langle X^i, X^j\rangle_t \\
&= Y_t \left(\sum_{i=1}^{n} \lambda_i\, dX_t^i + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j\, d\langle X^i, X^j\rangle_t\right) \\
&= Y_t \left(\sum_{i=1}^{n} \lambda_i\, dX_t^i + \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2\, dt\right),
\end{aligned}$$

where the last line follows from the fact that $\langle X^i, X^j\rangle_t = t$.

By integrating the equation for Y_t from 0 to t, and then taking expectations and using the fact that the stochastic integrals have zero expectation, we have

$$\mathbb{E}[Y_t] = 1 + \frac{1}{2}\sum_{i=1}^{n} \lambda_i^2 \int_0^t \mathbb{E}[Y_s]\, ds.$$

Solving this integral equation yields

$$\mathbb{E}\left[\exp\left(\sum_{i=1}^{n} \lambda_i X_t^i\right)\right] = \mathbb{E}[Y_t] = \exp\left(\frac{1}{2}\sum_{i=1}^{n} \lambda_i^2 t\right),$$

which shows that $\mathbf{X}_t \sim \mathcal{N}(0, tI_n)$. The independent, stationary, and Gaussian increment can then be shown by utilizing this finding and following the proof of Theorem 3.4.

□